

University of Groningen

Beyond OCR: Handwritten manuscript attribute understanding

He, Sheng

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version

Publisher's PDF, also known as Version of record

Publication date:

2017

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

He, S. (2017). *Beyond OCR: Handwritten manuscript attribute understanding*. [Thesis fully internal (DIV), University of Groningen]. University of Groningen.

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Beyond OCR: Handwritten Manuscript Attribute Understanding

Sheng He

Book cover is designed with help of Yanfang Feng to show the main idea of this thesis: training the computer to read the handwritten manuscript from the MPS data set and answer three questions: who wrote it, when and where was it written?

ISBN printed version: 978-90-367-9643-9

ISBN electronic version: 978-90-367-9642-2

Printed by: Ipskamp Drukkers, Enschede, The Netherlands.

Supported by the Netherlands Organisation for Scientific Research (NWO)
under project number 380-50-006



Netherlands Organisation for Scientific Research



university of
groningen

Beyond OCR: Handwritten Manuscript Attribute Understanding

PhD thesis

to obtain the degree of PhD at the
University of Groningen
on the authority of the
Rector Magnificus Prof. E. Sterken
and in accordance with
the decision by the College of Deans.
This thesis will be defended in public on
Friday 17 March 2017 at 12.45 hours

by

Sheng He
born on 1 July 1986
in Shaanxi, China

Supervisors

Prof. L.R.B. Schomaker

Prof. J.W.J. Burgers

Assessment committee

Prof. Cheng-Lin Liu

Prof. M. Biehl

Prof. E.O. Postma

Contents

1	Introduction	1
1.1	How to identify writers?	3
1.2	How to estimate date and geographical location?	5
1.3	Research questions	8
1.4	Material	10
1.5	Organization of this thesis	13
I	Writer Identification	15
2	Writer Identification Using Delta-n Hinge Feature	17
2.1	Introduction	17
2.2	Δ^n Hinge feature	18
2.3	Writer Identification	20
2.4	Experiments	21
2.5	Conclusion	25
3	Writer Identification Using Curvature-free Features	27
3.1	Introduction	27
3.2	Run-lengths of local binary pattern (LBPruns)	29
3.3	COLD feature	33
3.4	Experiments	36
3.5	Conclusion	46
4	Writer Identification Using Junction Features	47
4.1	Introduction	47
4.2	Related work	49
4.3	Junction detection	50
4.4	Writer identification	57
4.5	Experimental results	58
4.6	Conclusion	64

II	Historical document dating and localization	67
5	Historical Manuscript Dating Using Contour and Stroke Fragments.	69
5.1	Introduction	69
5.2	k Contour Fragments (kCF)	71
5.3	k Stroke Fragments (kSF)	75
5.4	Experiments	80
5.5	Discussion and conclusion	87
6	Historical Manuscript Dating and Localization Using A Multiple-Label Clustering Algorithm	89
6.1	Introduction	89
6.2	Histogram of Orientations of Handwritten Stroke Descriptor (H_2OS)	90
6.3	Multi-Label Self-Organizing Map (MLSOM)	93
6.4	Experiments	98
6.5	Conclusion	111
III	Critical Comparisons	113
7	Beyond OCR: Multi-faceted understanding of handwritten document characteristics	115
7.1	Introduction	115
7.2	Joint feature distribution principle	116
7.3	Feature representation	118
7.4	Applications	125
7.5	Discussion and conclusion	139
8	Discussion	141
8.1	Answers to the research questions	142
8.2	Future research	144
	Bibliography	149
	Summary	159
	Samenvatting	163
	Publications	167
	Acknowledgements	169